

Genome sequence of *Ensifer meliloti* strain WSM1022; a highly effective microsymbiont of the model legume *Medicago truncatula* A17

Jason Terpolilli¹, Yvette Hill¹, Rui Tian¹, John Howieson¹, Lambert Bräu², Lynne Goodwin³, James Han⁴, Konstantinos Liolios⁴, Marcel Huntemann⁴, Amrita Pati⁵, Tanja Woyke⁴, Konstantinos Mavromatis⁵, Victor Markowitz⁵, Natalia Ivanova³, Nikos Kyprides³, Wayne Reeve^{1*}.

¹ Centre for Rhizobium Studies, Murdoch University, Western Australia, Australia

² School of Life and Environmental Sciences, Deakin University, Victoria, Australia

³ Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

⁴ DOE Joint Genome Institute, Walnut Creek, California, USA

⁵ Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

*Correspondence: Wayne Reeve (W.Reeve@murdoch.edu.au)

Keywords: root-nodule bacteria, nitrogen fixation, rhizobia, *Alphaproteobacteria*

Ensifer meliloti WSM1022 is an aerobic, motile, Gram-negative, non-spore-forming rod that can exist as a soil saprophyte or as a legume microsymbiont of *Medicago*. WSM1022 was isolated in 1987 from a nodule recovered from the roots of the annual *Medicago orbicularis* growing on the Cyclades Island of Naxos in Greece. WSM1022 is highly effective at fixing nitrogen with *M. truncatula* and other annual species such as *M. tornata* and *M. littoralis* and is also highly effective with the perennial *M. sativa* (alfalfa or lucerne). In common with other characterized *E. meliloti* strains, WSM1022 will nodulate but fixes poorly with *M. polymorpha* and *M. sphaerocarpus* and does not nodulate *M. murex*. Here we describe the features of *E. meliloti* WSM1022, together with genome sequence information and its annotation. The 6,649,661 bp high-quality-draft genome is arranged into 121 scaffolds of 125 contigs containing 6,323 protein-coding genes and 75 RNA-only encoding genes, and is one of 100 rhizobial genomes sequenced as part of the DOE Joint Genome Institute 2010 Genomic Encyclopedia for Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB) project.

Introduction

An available source of nitrogen (N) is essential to life on Earth. Although the atmosphere consists of approximately 80% N, the overwhelming proportion of this is present in the form of dinitrogen (N₂) which is biologically inaccessible to the vast majority of higher organisms. Only a subset of microbes has the necessary molecular machinery to make atmospheric N₂ bioavailable by enzymatically reducing N₂ to NH₃. The fact that plant growth is most commonly limited by the availability of N may have provided the selective pressure for a wide range of plant genera, most of which are legumes, to evolve a symbiotic relationship with these N₂-fixing microbes. These microsymbionts, collectively termed root nodule bacteria, receive a carbon source from the plant and in return supply

the host with biologically fixed N. When these symbiotic interactions are optimally harnessed in agriculture, all the N-requirements of the host can be met, without the need to apply industrially synthesized N-based fertilizers, thereby increasing both the economic and environmental sustainability of the farming system [1].

Forage and fodder legumes play an integral role in sustainable farming practice, providing feed for stock while also enriching soil with bioavailable N. Worldwide, there are approximately 110 million ha of forage and fodder legumes under production [2], of which members of the *Medicago* genus comprise a considerable component. Two bacterial species, *Ensifer meliloti* and *E. medicae* are known to nodulate and fix N₂ with *Medicago* spp.

[3], although they differ in their symbiotic properties on some *Medicago* hosts. Specifically, while *E. medicae* can nodulate and fix N₂ with *M. murex*, *M. arabica* and *M. polymorpha*, *E. meliloti* does not nodulate *M. murex*, does not fix with *M. polymorpha* and fixes N₂ very poorly with *M. arabica* [4-6].

E. meliloti strain WSM1022 was isolated in 1987 from a nodule collected from the annual *M. orbicularis* growing on the Cyclades Island of Naxos in Greece. *E. meliloti* WSM1022 is a highly effective microsymbiont of *Medicago*, forming efficient N₂-fixing associations with the annual species *M. littoralis* and *M. tornata* [7]. In common with *E. medicae* WSM419 [8], WSM1022 also fixes approximately twice as much N₂ as *E. meliloti* 1021 on the model legume *M. truncatula* A17 [7]. However, unlike *E. medicae* WSM419, *E. meliloti* WSM1022 is also highly effective with the perennial *M. sativa* (alfalfa or lucerne) [7]. Therefore, *E. meliloti* WSM1022 is a broadly effective microsymbiont of *Medicago* spp. and as such represents a unique tool for the molecular analysis of effective N₂ fixation with fully sequenced macro- and microsymbionts. Here we present a summary classification and a set of general features for *E. meliloti* strain WSM1022 together with a description of its genome sequence and annotation.

Classification and features

E. meliloti WSM1022 is a motile, Gram-negative rod (Figure 1 Left and Center) in the order *Rhizobiales* of the class *Alphaproteobacteria*. It is fast growing, forming colonies within 3-4 days

when grown on half strength Lupin Agar (½LA) [9], tryptone-yeast extract agar (TY) [10] or a modified yeast-mannitol agar (YMA) [11] at 28°C. Colonies on ½LA are white-opaque, slightly domed and moderately mucoid with smooth margins (Figure 1Right).

Minimum Information about the Genome Sequence (MIGS) is provided in Table 1. Figure 2 shows the phylogenetic neighborhood of *E. meliloti* WSM1022 in a 16S rRNA sequence based tree. This strain shares 99.92% and 99.61% sequence identity (over 1290 bp) to the 16S rRNA of the fully sequenced *E. meliloti* 1021 [29] and *E. medicae* WSM419 [8] strains, respectively.

Symbiotaxonomy

E. meliloti strain WSM1022 was isolated in 1987 from a nodule collected from the annual *M. orbicularis* growing on the Cyclades Island of Naxos in Greece. The site of collection was a gentle slope and the soil a sandy-loam texture of pH 7.5-8.0. *E. meliloti* forms nodules (Nod⁺) and fixes N₂ (Fix⁺) on a range of annual *Medicago* spp. as well as the perennial *M. sativa* (Table 2). In common with other characterized *E. meliloti* strains, WSM1022 does not nodulate *M. murex*, does not fix N₂ with *M. polymorpha* and *M. arabica* [4,5] and is a poorly effective microsymbiont of *M. sphaerocarpus* [11]. However, WSM1022 is broadly effective with the alkaline soil-adapted annuals *M. littoralis* and *M. tornata* as well as the widely grown perennial forage legume *M. sativa*. In addition, WSM1022 is also a highly effective microsymbiont for the model legume *M. truncatula* A17.

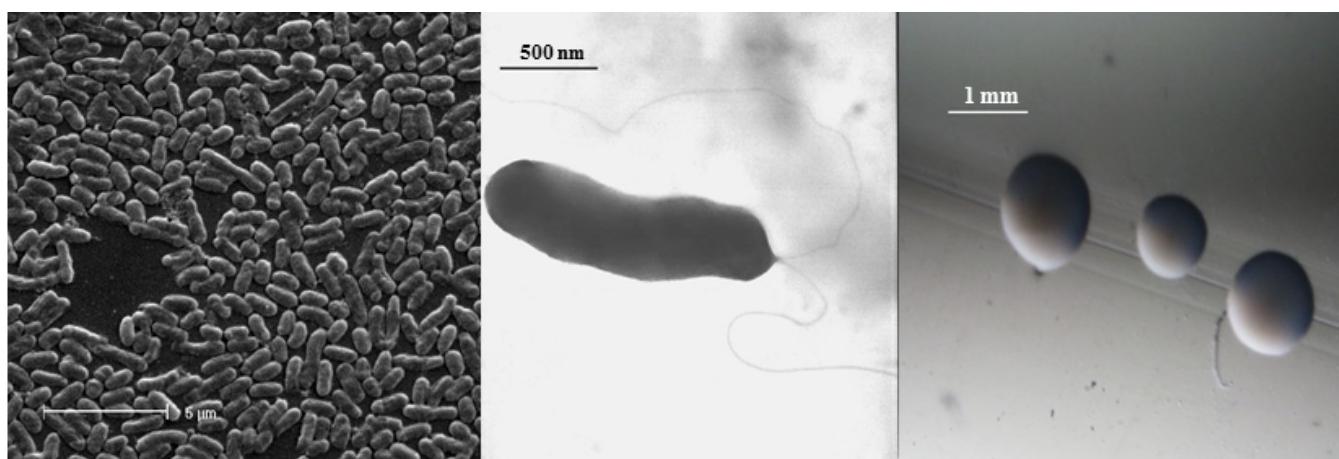


Figure 1. Images of *Ensifer meliloti* WSM1022 using scanning (Left) and transmission (Center) electron microscopy and the appearance of colony morphology on a solid medium (Right).

Table 1. Classification and general features of *Ensifer meliloti* WSM1022 according to the MIGS recommendations [12]

MIGS ID	Property	Term	Evidence code
Current classification		Domain <i>Bacteria</i>	TAS [13]
		Phylum <i>Proteobacteria</i>	TAS [14]
		Class <i>Alphaproteobacteria</i>	TAS [15,16]
		Order <i>Rhizobiales</i>	TAS [16,17]
		Family <i>Rhizobiaceae</i>	TAS [18,19]
		Genus <i>Ensifer</i>	TAS [20-22]
MIGS-22	Species <i>Ensifer meliloti</i>		TAS [21]
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Motile	IDA
	Sporulation	Non-sporulating	NAS
	Temperature range	Mesophile	NAS
	Optimum temperature	28°C	NAS
	Salinity	Non-halophile	NAS
MIGS-6	Oxygen requirement	Aerobic	TAS [7]
	Carbon source	Varied	NAS
	Energy source	Chemoorganotroph	NAS
MIGS-15	Habitat	Soil, root nodule, on host	TAS [7]
MIGS-14	Biotic relationship	Free living, symbiotic	TAS [7]
MIGS-4	Pathogenicity	Non-pathogenic	NAS
	Biosafety level	1	TAS [23]
	Isolation	Root nodule	TAS [11]
MIGS-5	Geographic location	Naxos, Greece	TAS [11]
MIGS-4.1	Soil collection date	28 April 1987	IDA
MIGS-4.2	Longitude	37.107772	IDA
MIGS-4.3	Latitude	25.387841	
MIGS-4.4	Depth	0-10cm	
	Altitude	Not recorded	

Evidence codes – IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [24].

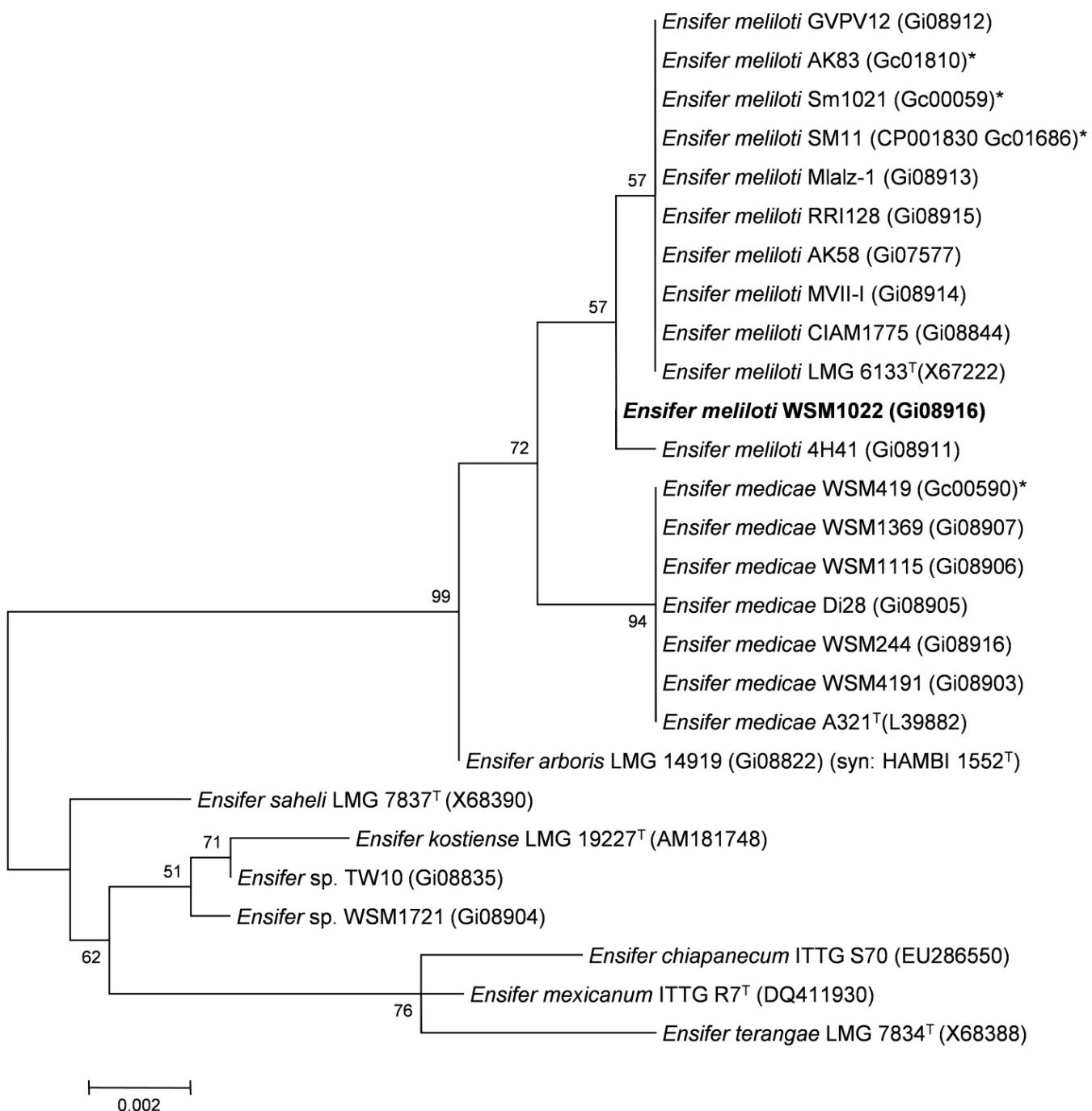


Figure 2. Phylogenetic tree showing the relationship of *Ensifer meliloti* WSM1022 (shown in bold print) to other *Ensifer* spp. in the order *Rhizobiales* based on aligned sequences of the 16S rRNA gene (1,290 bp internal region). All sites were informative and there were no gap-containing sites. Phylogenetic analyses were performed using MEGA, version 5 [25]. The tree was built using the Maximum-Likelihood method with the General Time Reversible model [26]. Bootstrap analysis [27] with 500 replicates was performed to assess the support of the clusters. Type strains are indicated with a superscript T. Brackets after the strain name contain a DNA database accession number and/or a GOLD ID (beginning with the prefix G) for a sequencing project registered in GOLD [28]. Published genomes are indicated with an asterisk.

Table 2. Nodulation and N₂ fixation properties of *E. meliloti* WSM1022 on selected *Medicago* spp.
Data compiled from [7,11][†]

Species Name	Cultivar or Accession	Growth Habit	Nodulation	N ₂ fixation	Comment
<i>M. truncatula</i>	A17	Annual	Nod ⁺	Fix ⁺	Highly effective
<i>M. truncatula</i>	Jemalong	Annual	Nod ⁺	Fix ⁺	Highly effective
<i>M. truncatula</i>	Caliph	Annual	Nod ⁺	Fix ⁺	Highly effective
<i>M. littoralis</i>	Harbinger	Annual	Nod ⁺	Fix ⁺	Highly effective
<i>M. tornata</i>	Tornafield	Annual	Nod ⁺	Fix ⁺	Highly effective
<i>M. sphaerocarpos</i>	Orion	Annual	Nod ⁺	Fix ⁺	Poorly effective
<i>M. arabica</i>	SA36043	Annual	Nod ⁺	Fix ⁻	No fixation
<i>M. polymorpha</i>	Santiago	Annual	Nod ⁺	Fix ⁻	No fixation
<i>M. murex</i>	Zodiac	Annual	Nod ⁻	Fix ⁻	No nodulation
<i>M. sativa</i>	Sceptre	Perennial	Nod ⁺	Fix ⁺	Highly effective

[†]Note that '+' and '-' denote presence or absence, respectively, of nodulation (Nod) or N₂ fixation (Fix).

Genome sequencing and annotation

Genome project history

This organism was selected for sequencing on the basis of its environmental and agricultural relevance to issues in global carbon cycling, alternative energy production, and biogeochemical importance, and is part of the Community Sequencing Program at the U.S. Department of Energy, Joint Genome Institute (JGI) for projects of relevance to agency missions. The genome project is deposited in the Genomes OnLine Database [28] and an improved-high-quality-draft genome sequence in IMG. Sequencing, finishing and annota-

tion were performed by the JGI. A summary of the project information is shown in Table 3.

Growth conditions and DNA isolation

E. meliloti WSM1022 was cultured to mid logarithmic phase in 60 ml of TY rich medium [30] on a gyratory shaker at 28°C. DNA was isolated from the cells using a CTAB (Cetyl trimethyl ammonium bromide) bacterial genomic DNA isolation method [31].

Table 3. Genome sequencing project information for *E. meliloti* WSM1022.

MIGS ID	Property	Term
MIGS-31	Finishing quality	Improved high-quality draft
MIGS-28	Libraries used	1× Illumina library
MIGS-29	Sequencing platforms	Illumina HiSeq 2000
MIGS-31.2	Sequencing coverage	Illumina: 275×
MIGS-30	Assemblers	Velvet version 1.1.04; Allpaths-LG version r42328
MIGS-32	Gene calling methods	Prodigal 1.4, GenePRIMP
	GOLD ID	Gi08916
	NCBI project ID	78233
	Database: IMG	2510065057
	Project relevance	Symbiotic N ₂ fixation, agriculture

Genome sequencing and assembly

The genome of *Ensifer meliloti* WSM1022 was sequenced at the Joint Genome Institute (JGI) using Illumina technology [32]. An Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform which generated 12,082,430 reads totaling 1812.4 Mbp.

All general aspects of library construction and sequencing performed at the JGI can be found at the JGI website [31]. All raw Illumina sequence data was passed through DUK, a filtering program developed at JGI, which removes known Illumina sequencing and library preparation artifacts (Mingkun, L., Copeland, A. and Han, J., unpublished). The following steps were then performed for assembly: (1) filtered Illumina reads were assembled using Velvet [33] (version 1.1.04), (2) 1–3 kb simulated paired end reads were created from Velvet contigs using wgsim (<https://github.com/lh3/wgsim>), (3) Illumina reads were assembled with simulated read pairs using Allpaths-LG [34] (version r42328). Parameters for assembly steps were: 1) Velvet (velveth: 63 –shortPaired and velvetg: –veryclean yes –exportFiltered yes –mincontiglngth 500 –scaffolding no–covcutoff 10) 2) wgsim (–e 0 –1 100 –2 100 –r 0 –R 0 –X 0) 3) Allpaths-LG (PrepareAllpathsInputs:PHRED64=1 PLOIDY=1 FRAGCOVERAGE=125 JUMPCOVERAGE=25 LONGJUMPCOV=50, RunAllpath-sLG: THREADS=8 RUN=stdshredpairs TARGETS=standard VAPIWARNONLY=True OVERWRITE=True). The final draft assembly contained 125 contigs in 121 scaffolds. The total size of the genome is 6.6 Mb and the final assembly is based on 1,812.4 Mbp of

Illumina data, which provides an average 275× coverage of the genome.

Genome annotation

Genes were identified using Prodigal [35] as part of the DOE-JGI annotation pipeline [36]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. The tRNAscanSE tool [37] was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA [38]. Other non-coding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERNAL (<http://infernal.janelia.org>). Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes (IMG-ER) platform [39].

Genome properties

The genome is 6,649,661 nucleotides with 62.16% GC content (Table 4) and comprised of 121 scaffolds (Figure 3) of 125 contigs. From a total of 6,398 genes, 6,323 were protein encoding and 75 RNA only encoding genes. The majority of genes (80.78%) were assigned a putative function whilst the remaining genes were annotated as hypothetical. The distribution of genes into COGs functional categories is presented in Table 5.

Table 4. Genome Statistics for *Ensifer meliloti* WSM1022

Attribute	Value	% of Total
Genome size (bp)	6,649,661	100.00
DNA coding region (bp)	5,733,017	86.22
DNA G+C content (bp)	4,133,661	62.16
Number of scaffolds	121	
Number of contigs	125	
Total gene	6,398	100.00
RNA genes	75	1.17
rRNA operons	1	0.02
Protein-coding genes	6,323	98.83
Genes with function prediction	5,168	80.78
Genes assigned to COGs	5,147	80.45
Genes assigned Pfam domains	5,331	83.32
Genes with signal peptides	563	8.80
Genes with transmembrane helices	1,437	22.93
CRISPR repeats	0	

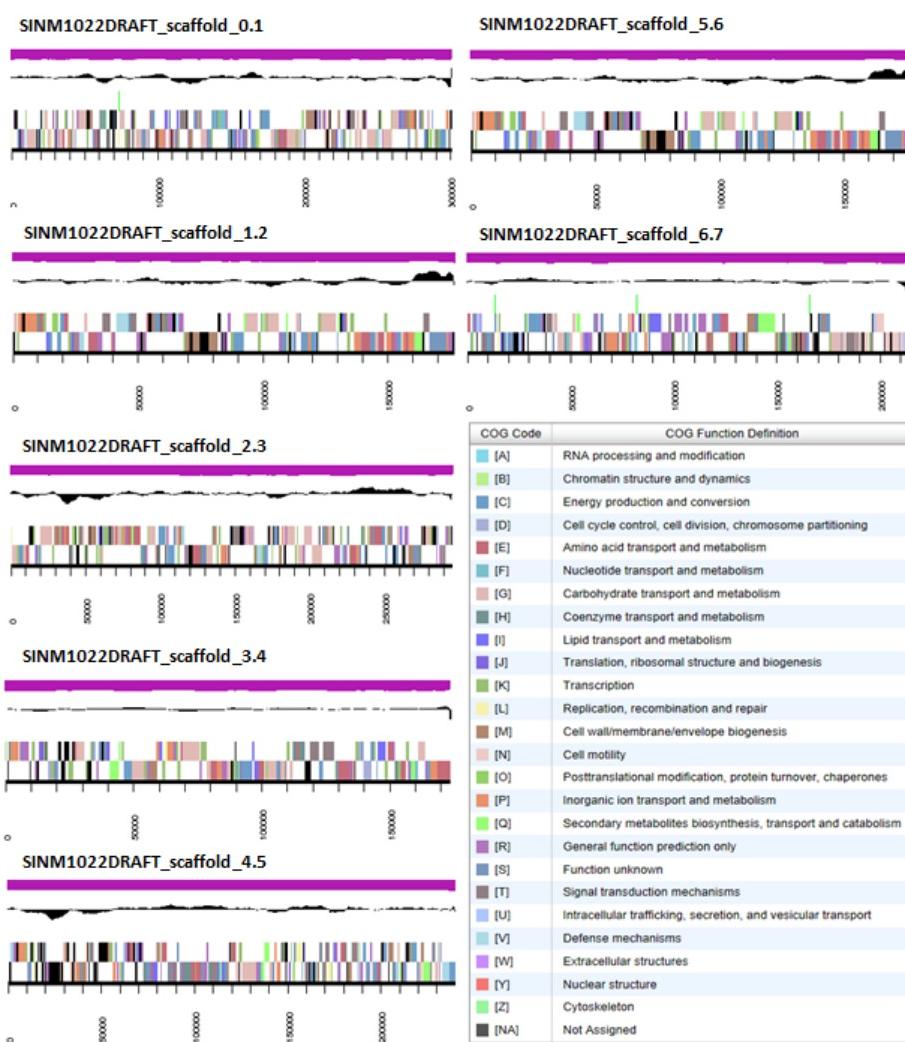
**Figure 3.** Graphical map of the genome of *Ensifer meliloti* WSM1022 showing the seven largest scaffolds. From bottom to the top of each scaffold: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.

Table 5. Number of protein coding genes of *Ensifer meliloti* WSM1022 associated with the general COG functional categories.

Code	Value	% age	COG Category
J	194	3.38	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	497	8.65	Transcription
L	196	3.41	Replication, recombination and repair
B	1	0.02	Chromatin structure and dynamics
D	38	0.66	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	61	1.06	Defence mechanisms
T	235	4.09	Signal transduction mechanisms
M	301	5.24	Cell wall/membrane biogenesis
N	71	1.24	Cell motility
Z	0	0.00	Cytoskeleton
W	1	0.02	Extracellular structures
U	113	1.97	Intracellular trafficking and secretion
O	177	3.08	Posttranslational modification, protein turnover, chaperones
C	357	6.21	Energy production conversion
G	606	10.54	Carbohydrate transport and metabolism
E	623	10.84	Amino acid transport metabolism
F	109	1.90	Nucleotide transport and metabolism
H	200	3.48	Coenzyme transport and metabolism
I	207	3.60	Lipid transport and metabolism
P	312	5.43	Inorganic ion transport and metabolism
Q	158	2.75	Secondary metabolite biosynthesis, transport and catabolism
R	708	12.32	General function prediction only
S	583	10.14	Function unknown
-	1,251	19.55	Not in COGS
Total	5,748	-	-

Acknowledgements

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396. We grate-

fully acknowledge the funding received from the Murdoch University Strategic Research Fund through the Crop and Plant Research Institute (CaPRI) and the Centre for Rhizobium Studies (CRS) at Murdoch University. We also acknowledge ECR funding for J. Terpolilli awarded by the School of Veterinary and Life Sciences at Murdoch University.

References

1. Howieson JG, O'Hara GW, Carr SJ. Changing roles for legumes in Mediterranean agriculture: developments from an Australian perspective. *Field Crops Res* 2000; **65**:107-122.
[http://dx.doi.org/10.1016/S0378-4290\(99\)00081-7](http://dx.doi.org/10.1016/S0378-4290(99)00081-7)
2. Herridge DF, Peoples MB, Boddey RM. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* 2008; **311**:1-18.
<http://dx.doi.org/10.1007/s11104-008-9668-3>
3. Graham P. Ecology of the root-nodule bacteria of legumes. In: Dilworth MJ, James EK, Sprent JI, Newton WE, editors. Nitrogen-Fixing Leguminous Symbioses. Dordrecht: The Netherlands: Springer; 2008. p 23-43.
4. Garau G, Reeve WG, Brau L, Yates RJ, James D, Tiwari R, O'Hara GW, Howieson JG. The symbiotic requirements of different *Medicago* spp. suggest the evolution of *Sinorhizobium meliloti* and *S. medicae* with hosts differentially adapted to soil pH. *Plant Soil* 2005; **276**:263-277.
<http://dx.doi.org/10.1007/s11104-005-0374-0>
5. Rome S, Cleyet-Marel JC, Matheron LA, Normand P, Brunel B. Rapid identification of *Medicago* nodulating strains by using two oligonucleotide probes complementary to 16S rDNA sequences. *Can J Microbiol* 1997; **43**:854-861. [PubMed](#)
<http://dx.doi.org/10.1139/m97-124>
6. Brunel B, Rome S, Ziani R, Cleyet-Marel JC. Comparison of nucleotide diversity and symbiotic properties of *Rhizobium meliloti* populations from annual *Medicago* species. *FEMS Microbiol Ecol* 1996; **19**:71-82. <http://dx.doi.org/10.1111/j.1574-6941.1996.tb00200.x>
7. Terpolilli JJ, O'Hara GW, Tiwari RP, Dilworth MJ, Howieson JG. The model legume *Medicago truncatula* A17 is poorly matched for N₂ fixation with the sequenced microsymbiont *Sinorhizobium meliloti* 1021. *New Phytol* 2008; **179**:62-66. [PubMed](#)
<http://dx.doi.org/10.1111/j.1469-8137.2008.02464.x>
8. Reeve W, Chain P, O'Hara G, Ardley J, Nandesena K, Brau L, Tiwari R, Malfatti S, Kiss H, Lapidus A, et al. Complete genome sequence of the *Medicago* microsymbiont *Ensifer* (*Sinorhizobium*) *medicae* strain WSM419. *Stand Genomic Sci* 2010; **2**:77-86. [PubMed](#)
<http://dx.doi.org/10.4056/sigs.43526>
9. Howieson JG, Ewing MA, D'antuono MF. Selection for acid tolerance in *Rhizobium meliloti*. *Plant Soil* 1988; **105**:179-188.
<http://dx.doi.org/10.1007/BF02376781>
10. Beringer JE. R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* 1974; **84**:188-198. [PubMed](#) <http://dx.doi.org/10.1099/00221287-84-1-188>
11. Terpolilli JJ. Why are the symbioses between some genotypes of *Sinorhizobium* and *Medicago* suboptimal for N₂ fixation? Perth: Murdoch University; 2009. 223 p.
12. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen M, Angiuoli SV, et al. Towards a richer description of our complete collection of genomes and metagenomes "Minimum Information about a Genome Sequence" (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed](#) <http://dx.doi.org/10.1038/nbt1360>
13. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990; **87**:4576-4579. [PubMed](#)
<http://dx.doi.org/10.1073/pnas.87.12.4576>
14. Garrity GM, Bell JA, Lilburn T. Phylum XIV. *Proteobacteria* phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 2, Part B, Springer, New York, 2005, p. 1.
15. Garrity GM, Bell JA, Lilburn T. Class I. *Alphaproteobacteria* class. In: Garrity GM, Brenner DJ, Kreig NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology. Second ed: New York: Springer - Verlag; 2005, p. 1.
16. Validation List No. 107. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* 2006; **56**:1-6. [PubMed](#) <http://dx.doi.org/10.1099/ijns.0.64188-0>
17. Kuykendall LD. Order VI. *Rhizobiales* ord. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), Bergey's Manual of Systematic Bacteriology, Second Edition, Volume 2, Part C, Springer, New York, 2005, p. 324.
18. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420.
<http://dx.doi.org/10.1099/00207713-30-1-225>
19. Conn HJ. Taxonomic relationships of certain non-sporeforming rods in soil. *J Bacteriol* 1938; **36**:320-321.
20. Casida LE. *Ensifer adhaerens* gen. nov., sp. nov.: a bacterial predator of bacteria in soil. *Int J Syst Bacteriol* 1982; **32**:339-345.
<http://dx.doi.org/10.1099/00207713-32-3-339>

21. Young JM. The genus name *Ensifer* Casida 1982 takes priority over *Sinorhizobium* Chen et al. 1988, and *Sinorhizobium morelense* Wang et al. 2002 is a later synonym of *Ensifer adhaerens* Casida 1982. Is the combination *Sinorhizobium adhaerens* (Casida 1982) Willems et al. 2003 legitimate? Request for an Opinion. *Int J Syst Evol Microbiol* 2003; **53**:2107-2110. [PubMed](#) <http://dx.doi.org/10.1099/ij.s.0.02665-0>
22. Judicial Commission of the International Committee on Systematics of Prokaryotes. The genus name *Sinorhizobium* Chen et al. 1988 is a later synonym of *Ensifer* Casida 1982 and is not conserved over the latter genus name, and the species name '*Sinorhizobium adhaerens*' is not validly published. Opinion 84. *Int J Syst Evol Microbiol* 2008; **58**:1973. [PubMed](#) <http://dx.doi.org/10.1099/ij.s.0.2008/005991-0>
23. Gubler M, Hennecke H, Fix A. B and C genes are essential for symbiotic and free-living, microaerobic nitrogen fixation. *FEBS Lett* 1986; **200**:186-192. [http://dx.doi.org/10.1016/0014-5793\(86\)80536-1](http://dx.doi.org/10.1016/0014-5793(86)80536-1)
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. [PubMed](#) <http://dx.doi.org/10.1038/75556>
25. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 2011; **28**:2731-2739. [PubMed](#) <http://dx.doi.org/10.1093/molbev/msr121>
26. Nei M, Kumar S. Molecular Evolution and Phylogenetics. New York: Oxford University Press; 2000.
27. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **39**:783-791. <http://dx.doi.org/10.2307/2408678>
28. Lioiios K, Mavromatis K, Tavernarakis N, Kyriakis NC. The Genomes On Line Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2008; **36**:D475-D479. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkm884>
29. Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, et al. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 2001; **293**:668-672. [PubMed](#) <http://dx.doi.org/10.1126/science.1060966>
30. Reeve WG, Tiwari RP, Worsley PS, Dilworth MJ, Glenn AR, Howieson JG. Constructs for insertional mutagenesis, transcriptional signal localization and gene regulation studies in root nodule and other bacteria. *Microbiology* 1999; **145**:1307-1316. [PubMed](#) <http://dx.doi.org/10.1099/13500872-145-6-1307>
31. DOE Joint Genome Institute user home. <http://my.jgi.doe.gov/general/index.html>
32. Bennett S, Solexa Ltd. *Pharmacogenomics* 2004; **5**:433-438. [PubMed](#) <http://dx.doi.org/10.1517/14622416.5.4.433>
33. Zerbino DR. Using the Velvet *de novo* assembler for short-read sequencing technologies. Current Protocols in Bioinformatics 2010; Chapter 11:Unit 11.5.
34. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA* 2011; **108**:1513-1518. [PubMed](#) <http://dx.doi.org/10.1073/pnas.1017351108>
35. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**:119. [PubMed](#) <http://dx.doi.org/10.1186/1471-2105-11-119>
36. Mavromatis K, Ivanova NN, Chen IM, Szeto E, Markowitz VM, Kyrpides NC. The DOE-JGI Standard operating procedure for the annotations of microbial genomes. *Stand Genomic Sci* 2009; **1**:63-67. [PubMed](#) <http://dx.doi.org/10.4056/sigs.632>
37. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. [PubMed](#)
38. Pruesse E, Quast C, Knittel K, Fuchs BdM, Ludwig W, Peplies J, Glöckner FO. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 2007; **35**:7188-7196. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkm864>
39. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. [PubMed](#) <http://dx.doi.org/10.1093/bioinformatics/btp393>